(1) Publication number:

0 079 739 **A2** 

(12)

# **EUROPEAN PATENT APPLICATION**

(21) Application number: 82305926.6

(22) Date of filing: 08.11.82

(5) Int. Cl.<sup>3</sup>: C 12 N 15/00 C 12 N 1/00, C 12 P 21/02 C 07 H 21/04, C 07 C 103/52 //C12R1/19, C12R1/865

30 Priority: 12.11.81 US 320632

(3) Date of publication of application: 25.05.83 Bulletin 83/21

(84) Designated Contracting States: BE CH DE FR GB IT LI NL SE 71) Applicant: THE UPJOHN COMPANY 301 Henrietta Street Kalamazoo, Michigan 49001(US)

(72) Inventor: Dugaiczyk, Achilles c/o The Upjohn Company 301 Henrietta Street Kalamazoo Michigan 49001(US)

(74) Representative: Perry, Robert Edward et al, GILL JENNINGS & EVERY 53-64 Chancery Lane London WC2A 1HN(GB)

(54) Albumin-based nucleotides, their replication and use, and plasmids for use therein.

(57) The DNA sequence coding for human serum albumin has been isolated and inserted as two fragments into two novel plasmids which can be replicated in E. coli. These novel fragments can be joined to provide a unitary DNA sequence which then can be cloned into a suitable host, e.g. E. coli, for the expression of human serum albumin (which is used extensively in medical practice in treating shock conditions).

GJE 70/2056/02

ALBUMIN-BASED NUCLEOTIDES, THEIR REPLICATION AND USE, AND PLASMIDS FOR USE THEREIN

This invention relates to nucleotides related to human serum albumin (HSA), their replication and use, and plasmids (and host substances) for use therein.

The gene for serum albumin is regulated in development. On the other hand, serum albumin is synthesis development by the adult liver, and its plateau in adulthood. The embryonic liver and yolk sac, on the other hand, produce predominantly  $\alpha$ -fetoprotein, but the synthesis decreases drastically after birth. Recently,

10 Law et al determined the complete sequence of mouse α-fetoprotein mRNA, Nature 291 (1981) 201-205. The structure revealed extensive homology to mammalian serum albumin, indicating that the two proteins are encoded in the same gene family. Similar conclusions have been

15 reached from studies on the α-fetoprotein genes of the rat and the mouse; see Jagodzinski et al, Proc. Natl. Acad. Sci. USA, 78 (1981) 3521-3525, and Gorin et al, J. Biol. Chem. 256 (1981) 1954-1959.

The complete nucleotide sequence of human serum 20 mRNA has been determined from recombinant cDNA clones and from a primer-extended cDNA synthesis on the mRNA 2,078 nucleotides, comprises template. The sequence starting upstream of a potential ribosome binding site in the 5'-untranslated region. It contains all the 25 translated codons and extends into the poly(A) at the 3'-terminus. Part of the translated sequence codes for a hydrophobic prepeptide met-lys-trp-val-thr-phe-ile-serleu-leu-phe-leu-phe-ser-ser-ala-tyr-ser, followed by a basic propeptide arg-gly-val-phe-arg-arg. These signal 30 peptides are absent from mature serum albumin and, so far, have not been identified in their nascent state in humans. A remaining 1,755 nucleotides of the translated mRNA sequence code for 585 amino acids which are in agreement, with few exceptions, with the published amino 35 acid data for human serum albumin. The mRNA sequence

verifies and refines the repeating homology in the triple-

domain structure of the serum albumin molecule.

## DETAILED DESCRIPTION OF THE INVENTION-

Human serum albumin cDNA is cloned into the PstI site of plasmid pBR322 by the oligo(dG)-oligo(dC) tailing technique. Plasmid DNA was isolated from 97 positive colonies which hybridized to the enriched albumin cDNA probe, and the recombinant plasmid pHA36 was found to contain the largest insert of an albumin cDNA sequence. Its restriction endonuclease map is shown in the drawing, together with a restriction map of the primer-extended plasmid clone pHA206. The latter was obtained in a second transformation experiment after initiating the cDNA synthesis from an internal primer. This primer was a 91 base pairs long DNA fragment, MspI(152)-TaqI(182/3), isolated from pHA36. The two plasmids, pHA36 and pHA206, share 0.15 kb of homologous DNA. Together, they encode the entire sequence for human serum albumin, starting with the CTT codon for leu -10 of the prepeptide and extending into the 3'-untranslated region of poly(A).

Sequence of the Albumin cDNA. The sequence was determined for the most part on both DNA strands to ensure accuracy. All of the restriction sites used to end-label DNA fragments were sequenced across by 20 labeling a neighboring restriction site. The entire nucleotide sequence of the serum albumin mRNA, as determined from the cloned DNA in pHA36, pHA206, and from the primer-extended cDNA at the 5'-terminus of the message, is shown in the following Table 1. The inferred amino acid sequence is also indicated. The mRNA length is 2,078 nucleo-<sup>25</sup> tides, of which 38 represent the 5'-untranslated region, 54 identify a prepeptide of 18 amino acids, 18 identify a propeptide of 6 amino acids, 1,755 code for the known 585 amino acids of serum albumin, 189 make up the 3'-untranslated region and 24 are the poly(A) sequence. Nucleotides 5 to 15 (-34 to -24) in the 5'-untranslated region (Table  $^{30}$  1) are complementary to a 3'-terminal region of eukaryotic 18S RNA [Azad, A.A. and Deacon, N.J. (1980) Nucl. Acids Res. 8, 4365-4376] and thus could represent a ribosome binding site:

(5')...T 
$$T^CT$$
 C T T C T G T.....albumin mRNA (3')...G A G G A A G G C G U C C  $m_2^6$ A  $m_2^6$ A.....185 RNA

The translated portion of the mRNA sequence codes for the signal peptide and the main body of the albumin polypeptide chain. The

4083

signal peptide is composed of a hydrophobic prepeptide of 18 amino acids and a basic propeptide of 6 amino acids (Table 1). Since prepeptides are removed from nascent secretory proteins (like albumin) in the endoplasmic reticulum, they are seen only in vitro in heterologous translation systems. As yet, they have not been found within cells [Judah, J.D. and Quinn, P.S. (1977) FEBS 11th Mtg., Copenhagen 50, 21-29; and Strauss, A.W., Donohue, A.M., Bennett, C.D., Rodkey, J.A. and Alberts, A.W. (1977) Proc. Natl. Acad. Sci. USA 74, 1358-1362]. This is the first report of the presence and the sequence of a prepeptide for human serum albumin. As it is with other secretory proteins, the conversion of proalbumin to albumin takes place in the Golgi vesicles, and the enzyme responsible for this cleavage is probably cathepsin B [Judah, J.D. and Quinn, P.S. (1978) Nature 271, 384-385]. This is also a first report on the sequence of the propeptide for normal human serum albumin.

At the 3'-end of the message, the putative polyadenylation signal sequence, AATAAA, is located 164 nucleotides downstream from the amino acid termination codon TAA and 16 nucleotides upstream from the beginning of the poly(A) sequence. Another characteristic sequence located near the polyadenylation site has been identified by Benoist, et al. [Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Nucl. Acids Res. 8, 127-142]; the concensus sequence from several mRNAs was concluded as TTTTCACTGC. A similar sequence, TTTTCTCTGT, is located 19 nucleotides upstream from the AATAAA hexanucleotide in the human albumin mRNA (Table 1).

# TABLE 1

				IAB	LE I				
	(30)	(170)	(560)	(350)	(440)	(330)	(029)	(710)	(300)
5	Ser AGC	20 1ys AAA	50 a1a GCA	80 leu CTT	110 pro	140 try TAT	170 q1n CAA	200 cys TCT	230 q1u GAA
3		phe TTC	phe TTT		asn AAC			1ys AAG	ala GCA
	leu phe CTC TTT	asn phe AAT TTC	a S	ala thr GCA ACT	asp asn GAC AAC	tyr TAC	168 cys TGT	leu CTC	phe TTT
	phe	1₽ <b>6</b> A	thr	val	asp	130 asn glu glu thr phe leu lys lys tyr leu AAT GAA GAG ACA TTT TTG AAA AAA TAC TTA	glu	ala lys qin arq leu lys GCC AAA CAG AGA CTC AAG	pro lys ala glu CCC AAA GCT GAG
	18 CH	ala GA	val	thr	lys asp AAA GAT	1,5 <b>A A</b>	th ACA	gln CAG	ala CCT
10	-10 leu leu phe CTT CTT TTT	leu aly alu alu TTG GGA GAA GAA	glu val GAA GTA	75 leu cys thr TTA TGC ACA	hts CAC	leu TTG	ala phe GCT TTT	lys AA	pro lys CCC AAA
	ser TCC	leu aly TTG GGA	AAT		phe leu gin his lys TTC TTG CAA CAC AAA	thr phe leu lys lys ACA TTT TTG AAA AAA		31a	
	11e ATT	asp	Jeu val TTA GTG	asp 1ys GAC AAA	1ec 11G	th ACA	160 arg tyr lys ala AGG TAT AAA GCT	ser	220 ser gin arg phe AGC CAG AGA TTT
	phe TTT	lys AA	Jeu TT	asp GAC	phe TTC	d Ju	1ys AAA	ser	arg AGA
	p r o trp val tlu phe TGG GTA ACC TTT	10 arg phe CGG TTT	40 val lys gta aaa	91,	100 101 glu cys GAA TGC	glu GAA	tyr TAT	ala GCT	e de CYC
15	p r val t GTA A		40 481 GTA	70 phe			160 arg AGG	190 glu leu arg asp qlu qly lys GAA CTT CGG GAT GAA GGG AAG	
	lys trp AAG TGG	ala his GCT CAT	qlu asp hfs GAA GAT CAT	thr leu ACC CTT	asn	thr ala phe his asp ACT GCT TTT CAT GAC	ohe ala lys TTT GCT AAA	41. 666	arg leu CGC CTG
	1ys	ala GCT	asp GAT	leu his thr CTT CAT ACC	glu pro aly ard GAA CCT GGG AGA	h is	ala CCT		val ala arg GTA GCT CGC
•	-18 Het TCTTTCTGTCAACCCCACAGCCTTTGGCACA ATG	val	a) c GAA	ser leu his TCA CTT CAT	gin glu pro aly CAA GAA CCT GGG	ag E	glu leu leu phe phe GAA CTC CTT TTC TTT	asp	ala GCT
20	CACA	n té	phe	CTT	pro CCT	thr ala ACT GCT	phe TTC	918 000	ala trp ala val GCA TGG GCA GTA
	וופמ	ser AGT	PTO	lys ser AAA TCA	₹8	thr	leu CTT	les CTT	င် နေ
	מככד	lys AAG	34 gln cys pro CAG TGT CCA	1 ys	g]a	124 cys 7GC	leu CTC	glu GAA	£ 55
	ACAG	ala his lys ser GCA CAC AAG AGT	919 CAG	asp	1ys AAA	met ATG	al S	180 leu leu pro lys leu asp CTG TTG CCA AAG CTC GAT	phe lys ala trp ala TTC AAA GCA TGG GCA
	ວັນວ	ala GCA	gln CAG	62 cys 1GT	ala GCA	val	570	leu CTC	¥ ¥8
25	ICA	asp GAT	leu CTT	60 glu asn GAA AAT	24 161	asp	200 200	1ys AG	210 ala phe GCT TTC
	TCTG	pro-1 gly val phe arg arg GGT GTG TTT CGT CGA	30 tyr TAT	9 to	20 273 160	120 glu val GAG GTT	150 tyr TAT	180 Pro CCA	210 ala
	CTCT	arg CGT	gln	ala GCT	asp GAC	g GAG	phe TTT	1eu 176	P AG
	CTTT	eg E	ala GCT	ser TCA	ala GCT	CCA CCA	tyr TAC	leu CTG	88
	ઇ	p val GTG	phe TTT	91°	glu met GAA ATG	arg AGA	his pro CAT CCT	177 ala cys GCC TGC	9. 86.
30		gly GGT	ala တင	asp GAT		arg leu val arg CGA TTG GTG AGA		200 .	phe III
		arg AGG	11e ATT	ala GCT	tyr aly TAT GGT	arg leu CGA TTG	arg arg AGA AGA	ala CCT	17.
		-1 ser TCC	15	val CTT		arg CGA	Ag S	AAA	2
		ser ala tyr ICG GCT IAT	leu val leu lle TTG GTG TTG ATT	cys TGT	glu thr GAA ACC	leu pro CTC CCC	ata CCC	GAT	210 ser leu gin lys phe gly glu arg ala AGT CTC CAA AAA ITT GGA GAA AGA GCT
35		ala GCT	1eu 11G	thr			11e	ala	AGT
		3er 100	21 ala GCC	22 8	81 arg CGT	111 asn AAC	18 g g g g g g g g g g g g g g g g g g g	171 ala GCT	201 ala GCC

	(890)	(980)	(1070)	(1160)	(1250)	(1340)	(1430)	(1520)	(1610)	(1700)
_	260 lev CTT	290 11e ATT	320 ala GrT	350 ala GCC	380 Jeu CCT	410 arg CGT	440 his cat	470 ser AGT	500 17s AAA	530 val GTT
5	asp GAC	289 cys TGC	tyr TAT	leu CTT	PTO CCT	val GTT	1ys AAA	val CTA	ورر	leu CTT
	3)3 CCC	his CAC	asn AAC	ara leu AGA CTT	1ys AAA	leu TTA	438 cys TGT	PTO CCA		ala GCA
	arg AGG	ser TCC		val leu leu leu arg GTG CTG CTG AGA	alu phe GAA TTT	ala leu leu val RCG CTG TTA GTT	437 cys TGT		tyr TAC	
	asp arg GAC AGG	1ys AAA	316 cys TGC	leu CTG	olu GAA		1ys AAA	1 ys		gln thr CAA ACT
10	ala asp CCT CAT	lev glu TTG CAA	asp val	val leu GTC CTC	asp GAT	qln asn CAG AAT	aly ser CCC ACC	atu GAG	9. CA	lys lys AAG AAA
	ala asp CCT CAT	Jeu TTG	asp		phe TTC	4 p	ور ورو	CAT		1ys AAG
	253 glu cys GAA TGT	pro leu CCT CTG	1ys AAG	ser väl TCT GTC	lys val AAA GTG	lys phe AAA TTC	lys val AAA GTG	Jeu 116	glu val GAA GTC	520 glu arg gln fle GAG AGA CAA ATC
	253 glu cys GAA TGT		Ser	3er 701	1ys AAA	1ys AAA	<del>2,</del> ₹	val	91u 6AA	aln CAA
	leu CTT	280 qlu lys GAA AAA	g to	tyr TAC	370 tyr ala TAT GCC	tyr TAC	430 lev qly CTA GGA	460 461 leu cys TTA TGT	Jeu CTG	520 glu arg ain ile GAG AGA CAA ATC
15	250 Jeu CTG	280 q1u GAA	310 val CTT	340 asp GAT		400 qlu tyr GAG TAC		460 Jeu TTA		520 qlu GAG
	asp	279 cys 1GT	asp phe CAT TTT	pro CCT	369 atu cys GAA TGC	leu aly CTT GGA	AAC	gln CAG	phe ser TTT TCA	qlu lys GAG AAG
	÷ 5	278 279 qlu cys cys GAA TGC TGT	asp GAT	h1s CAT	a]u GAA	alu aln leu aly GAG CAG CTT GGA	arg AGA	ser val val 1eu asn TCC GTG GTC CTG AAC		ser qlu lys TCT GAG AAG
	his CAT		ala GCT	glu tyr ala arg arg GAA TAT GCA AGA AGG	\$ \$	91° CAG	ser TCA	Jeu CTG	pro cys CCA TGC	514 11e cys thr leu ser ATA TGC ACA CTT TCT
20	246 cys TGC	lys leu lys AAA CTG AAG	ser leu ala TCA TTA GCT	ara AGA	pro		glu val GAG GTC	دع 10	5 P P P P P P P P P P P P P P P P P P P	thr leu ACA CTT
20	245 cys TGC	Jeu CTG	ser leu TCA TTA	ala GCA	asp	phe TT		val GTG	\$1.5 CCA	thr ACA
		lys AA	ser TCA	tyr TAT	ele CA	glu Jeu GAG CTT	420 thr pro thr leu val ACT CCA ACT CTT GTA	1 S	480 ser leu val asn arg TCC TTG GTG AAC AGG	514 11e cys ATA TGC
	thr	ser AGT	pro CCT	91u GAA	ala GCT		leu CTT	tyr leu IAT CTA	asn	11e ATA
		ser TCC	leu TTG	tyr TAT	ala SCC	392 cys TGT	thr		val GTG	asp
25	240 1ys val AAA GTC	11e ATC	asp	330 phe leu tyr TTT TTG TAT	361 cys TGT	asn	PT CCA	450 glu asp GAA GAC	leu TTG	ala GCA
		270 ser TCG	300 ala GCT	330 met phe leu ATG TTT TTG	360 cys TGC	390 91n CAA	420 thr ACT	450 g1u GAA	480 <b>Ser</b> 100	510 phe h1s TTC CAT
	thr	asp	pro CCT	met ATG	glu lys GAG AAG	lys AAA	ser	ala GCA	glu GAA	
	asp leu GAT CTT	gla	met ATG	91y GGC	glu GAG	11e ATC	val	448 cys TGT	th ACA	thr
		glu asn GAA AAT	asp glu met GAT GAG ATG	phe leu gly r TTC TTG GGC /	leu CTA	leu TTA	g1 CAA	979 CC	476 477 cys cys TGC TGC	phe TTC
30	thr		asp	phe TTC	thr	asn	5 S	met ATG		thr
	val GTG	265 11e cys ATC TGT	glu asn GAA AAT	val	glu thr GAA ACC	olo CAG	val GTA	arg AGA	thr 1ys ACC AAA	9F 6A
	leu TTA	11e ATC	glu GAA	asp	glu GA	pro CCT	1ys AAA	lys AA		ala GCT
	ser lys leu val thr asp TCC AGG TTA GTG ACA GAT	tyr 1AT	val GTG	lys AG	tyr TAT	glu pro gin asn leu ile GAG CCT CAG AAT TTA ATC	1ys AAG	al a	arg val AGA GTC	asn
	ser	1ys AAG	g lu GAA	ala GCA	th ACA	ale GAA	thr			501 giu phe asn ala giu thr phe thr CAG TIT AAT GCT GAA ACA TTC ACC
35	231 val GTT	261 ala GCC	291 ala GCC	321 glu GAG	351 1ys AAG	381 val GTG	411 tyr TAC	441 pro CCT	471 asp GAC	501 91u GAG

	(1790)	(1883)	ter ter Catctcagcctaccatgagaataagagaaaaaatgaagatcaaaagcttattcatetettttegttiggtgtaaagccaacacctgtctaaaaaacataaatttetttaa (2002)
	560 175 AAG	<b>9</b> ≯	
5	559 5 cys 1 TGC /	ter TAN	CTT
Э	558 5 cys c TGC 1	CATI	ATT
	ys c AG 1	, ATCA	AT A
	lo J	ئ چ ۂ	AAAC
	6 S	2 ₹	Z Z Z
10	ÿ 5 • ⊢	~ F ~ y	:CTC
10	a p T	- 00 - ×	1222
	اء عاد	s le	Y)
	າຄ ລ	a a 1 GC	۷وږ
	d T	C a C	STAA
15	550 asp	580 r g n	<u>7667</u>
	£ 3	AGI	rceri
	ATC	£ 33	
	550 lys ala thr lys glu gin leu lys ala val met äsp asp phe ala ala phe val glu lys cys AAG GCA ACA AAA GAG CAA CTG AAA GCT GTT ATG GAT GTT TTC GCT GCT TTT GTA GAG AAG TGC	570 phe ala glu glu gly lys lys leu val ala ala ser gln ala ala leu gly leu ter 'ter TTT GCC GAG GAT AAA AAA CTT GTT GCT GCA AGT CAA GCT GCC TTA GGC TTA TAA CATCACATTTAAAAG (	110
	5 5 6 19	ral CTI	CTT)
20		CTT	ATCT
	leu CTG	1 ys	ATTC
	cAA	1ys AA	C11
	g Ja GAG	41 y	¥¥
	Lys AA	95	GATC
25	Sto thr	570 glu GAG	ICAA
	هاه 20	နှိုန ၁၁၁	¥
	1ys AAG	phe TTT	¥
	973 CCC	567 cys TGC	5
	lys AAG	glu thr GAG ACC	ter ter ITGAGAATAA
30	h1s CAC	alu GAG	SG C
•	leu val lys his lys CTC GTG AAA CAC AAG	asp lys c	CCA1
	val GTG	asp	CCTA
	leu CTC	asp	TCAG
35	531 glu GAG	561 ala GCT	CATC
35			_

TCATTTTGCCTCTTTTCTCTGTGCTTCAATTAATAAAATGGAAAGAATCTAA.... 20 ....AA (2078)

Following are examples which illustrate procedures, including the best mode, for practicing the invention. These examples should not be construed as limiting. All percentages are by weight and all solvent mixture proportions are by volume unless otherwise noted.

### Example 1 Isolation of Messenger RNA

Human liver mRNA was obtained following the procedure of Chirgwin, et al [Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294-5299]. Immunoprecipitation of albumin containing polysomes was performed according to Taylor and 10 Tse [Taylor, J.M. and Tse, T.P.H. (1976) J. Biol. Chem. 251, 7461-7467]. In vitro translation of mRNA was carried out in a reticulocyte cell-free system, following the instruction of the manufacturer (New England Nuclear). The translation products were separated electrophoretically according to Laemmli [Laemmli, J.K. (1970) Nature 227, 680-685.

#### Example 2 Cloning Procedures

30

Double stranded cDNA was synthesized as described previously [Law. S., Tamaoki, T., Kreuzaler, F. and Dugaiczyk, A. (1980) Gene 10, 53-61]. It was annealed to PstI-linearized pBR322 DNA [Rolivar, F., 20 Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L., Boyer, H.W., Crossa, J.H. and Falkow, S. (1977) Gene 2, 95-113] that had been tailed with 15 dG residues/3'-terminus [Dugaiczyk, A., Robberson, D.L. and Ullrich, A. (1980) Biochemistry 19, 5869-5873]. The annealed DNA was used to transform E. coli strain RR1, as detailed previously [Law, 25 S., et al., <a>Ibid.</a>]. The albumin clones were selected using the colony hybridization method of Grunstein and Hogness [Grunstein, M. and Hogness, D.S. (1975) Proc. Natl. Acad. Sci. USA 72, 3961-3965], with [<sup>32p</sup>]-labeled cDNA synthesized with the immunoprecipitated polysomal mRNA as template.

As shown in Example 5, plasmids pHA36 and pHA206 were deposited in E. coli HB101 hosts. The plasmids were obtained from E. coli RR1 hosts, described in this example, and transformed into E. coli HR101 by standard procedures well known to those of ordinary skill in this The E. coli RR1 hosts were lysed and then centrifuged to 35 separate the chromosomal DNA, cell DNA and plasmid DNA. The plasmid DNA, remaining in the supernatant, is precipitated with ethanol and the precipitate is resuspended in buffer, e.g., TCM (10mM Tris·HCl, pH 8.0, 10 mM CaCl<sub>2</sub>, 10 mM MgCl<sub>2</sub>). The cells for transformation are prepared as follows: 120 ml of L-broth (1% tryptone, 0.5% yeast extract, 0.5% NaCl) are inoculated with an 18 hour culture of HR101 NRRL B-11371 and grown to an optical density of 0.6 at 600 nm. Cells are washed in cold 100 mM NaCl and resuspended for 15 minutes in 20 ml chilled 50 mM CaCl<sub>2</sub>. Bacteria are then concentrated to one-tenth of this volume in CaCl<sub>2</sub> and mixed 2:1 (v:v) with annealed plasmid DNA, prepared as described above. After chilling the cell-DNA mixture for 15 minutes, it is heat shocked at 42°C for 2 minutes, then allowed to equilibrate at room temperature for ten minutes before addition of L-broth 10 times the volume of the cell-DNA suspension. Transformed cells are incubated in broth at 37°C for one hour before inoculating selective media (L-agar plus 10 µg/ml tetracycline) with 200 µl/plate. Plates are incubated at 37°C for 48 hours to allow the growth of transformants.

# 15 Example 3 Mapping of Restriction Endonuclease Sites

Restriction endonucleases were obtained from Bethesda Research Laboratories and New England Biolabs and were used according to the manufacturers' instructions. The digested DNA fragments were analyzed electrophoretically on agarose [Helling, R.B., Goodman, H.M. and Boyer, H.W. (1974) J. Virol. 14, 1235-1244] or acrylamide [Dingman, C., Fisher, M.P. and Kakefuda, T. (1972) Biochemistry 11, 1242-1250] gels.

## Example 4 DNA Sequencing

DNA fragments were dephosphorylated with bacterial alkaline phosphatase (Worthington) and labeled at the 5'-ends with polynucleotide kinase (Boehringer-Mannheim) and \(\gamma\big[^{32p}\Barparage]\)ATP. Following digestion with a second restriction endonuclease and electrophoretic separation of the fragments, DNA sequence determination was done according to the procedure of Maxam and Gilbert [Maxam, A. and Gilbert, W. (1980) Methods Enzym. 65, 499-560] and the degradation products were separated electrophoretically on 0.4 mm acrylamide gels as described by Sanger and Coulson [Sanger, F. and Coulson, R. (1978) FEBS Letters 87, 107-110].

# Example 5 Recombinant Plasmids pHA36 and pHA206

As disclosed in Example 2, albumin clones were selected by hybridizing to the enriched albumin cDNA probe. Plasmid pHA36 contained the largest insert of an albumin cDNA sequence. Both plasmids pHA36 and pHA206 have been deposited in a viable E. coli host in the

permanent collection of the Northern Regional Research Laboratory (NRRL), U.S. Department of Agriculture, Peoria, Illinois, U.S.A. Their accession numbers in this repository are as follows:

HB101(pHA36) - NRRL B-12551

5

HB101(pHA206) - NRRL B-12550

 $\underline{\text{E. coli}}$  HB101 is a known and widely available host microbe. Its NRRL accession number is NRRL B-11371.

NRRL B-12550 and NRRL B-12551 are available to the public. upon the grant of a patent. It should be understood that the availability of these deposits does not constitute a license to practice the subject invention in derogation of patent rights granted with the subject instrument by governmental action.

<u>E. coli</u> RR1 and <u>E. coli</u> HB101 are known and widely available host microbes. Their NRRL accession numbers are NRRL B-12186 and NRRL B-11371, respectively.

pBR322 is a well known and widely available plasmid. It can be obtained from the following host deposit by standard procedures:

NRRL B-12014 - E. coli RR1 (pBR322).

YEp6 is a well known and widely available yeast episomal plasmid.

20 It can be obtained from the following host deposit by standard procedures:

E. coli HB101 (YEp6) - NRRL B-12093.

# Example 6 Assembly of the Serum Albumin Gene

Assembling the pieces together is a straighforward task of restriction enzymology. There is only one MspI site in the overlapping
DNA sequence of the two cDNA clones. Two enzymatic steps of (i) MspI
digestion of the two DNAs, followed by (ii) the use of ligase, an
enzyme that seals DNA fragments, will give the desired product.
Although two other undesired DNA species will also be obtained in the
course of this recombination reaction, both of them will differ substantially in size. Thus, separation and isolation of the desired DNA
species will be achieved.

The assembled DNA clone can be used to transform two types of cells:

(a) Escherichia coli

- (b) Saccharomyces cerevisiae
- (a) The vector of choice is plasmid pRR322, the same that has

been successfully used for cloning of the two fragmented pieces of the serum albumin cDNA.

(b) In order to transform yeast with the serum albumin structural gene sequence, the DNA must be inserted into one of the existing yeast plasmid vectors. This can be accomplished by taking advantage of the fact that several restriction endonuclease recognition sequences are absent from the cloned serum albumin DNA. Synthetic <u>EcoRl</u> DNA linkers can be ligated to the DNA fragment containing the serum albumin sequence followed by insertion (ligation) into one of the yeast plasmid vectors, e.g., YEp6, at the <u>Eco</u> Rl cloning site. The fused chimeric plasmid can be used to transform yeast according to an established procedure [Hinnen, A., Hicks, J.B. and Fink, G.R. (1978) <u>Proc. Natl. Acad. Sci. USA</u>, 75, 1929]. YEp6 can be obtained from the NRRL repository, as disclosed supra.

# 15 Example 7 Expression of the Serum Albumin Gene

The main body of the structural gene will be transcribed by the E. coli or yeast enzymes. If little or no albumin is produced with the selected host, then an Escherichia coli promoter DNA sequence carrying an initiation codon, i.e., ATG, can be ligated at the begin-20 ning of the serum albumin structural gene. Such elements are known and available, e.g., lac promoter used for the expression of human interferon gene in E. coli [Proc. Natl. Acad. Sci. 77, 5230 (1980)]; source of promoter DNA [Proc. Natl. Acad. Sci. 76, 760 (1979)]. Also, see Nature, Vol. 281, October 18, 1979. It has already been documented that such Escherichia coli promoter sequences function well in the expression of foreign genes in Escherichia coli [Mercereau-Puijalon, O., Royal, A., Cami, B., Garapin, A., Krust, A., Gannon, I. and Kourilsky, P. (1978) Nature 275, 505; and Goeddel, D.V., Kleid, D.G., Bolivar, F., Heyneker, H.L., Yansura, D.G., Grea, R., Hirose, 30 T., Kraszewski, A., Itakura, K., and Riggs, A. (1979) Natl. Acad. Sci. USA 76, 106]. For expression in yeast, see Rose, M., Casadaban, M.J. and Botstein, D. (1981) Proc. Natl. Acad. Sci. USA 78, 2460 and 4466. Example 8 Screening of Clones Producing Albumin

Immunological methods can be used to detect small amounts of albumin made in a bacterium. Flat disks of flexible polyvinyl are coated with the IgG fraction from an immune serum and the disks are pressed onto an agar plate so that antigen released from an <u>in situ</u> lysed microbial colony can bind to the fixed antibody. The plastic

disk is then incubated with the same total IgG fraction labeled with radioactive iodine so that other determinants on the bound antigen can in turn bind the iodinated antibody. Radioactive areas on the disk expose X-ray film during autoradiography and thus identify colonies producing the protein which is being screened for. Detailed protocols of this procedure have been published [Broome, S. and Gilbert, W. (1978) Proc. Natl. Acad. Sci. USA, 75, 2746]. The purification of human serum albumin can be accomplished by using procedures well known in the art. For example, procedures disclosed in a chapter by T. Peters: Purification and Properties of Serum Albumin, in: The Plasma Proteins, Putnam, Ed. Academic Press, New York, 1975, can be used.

The work described herein was all done in conformity with physical and biological containment requirements specified in the NIH Guidelines.

# CLAIMS

- 1. Plasmid pHA36, having a restriction endonuclease pattern as shown in the drawing.
- 2. Plasmid pHA206, having a restriction endonuclease pattern as shown in the drawing.
- 3. E. coli HB101 (pHA36) having the deposit accession number 10 NRRL B-12551.
  - 4. E. coli HB101 (pHA206) having the deposit accession number NRRL B-12550.
- 5. A microorganism modified to contain a nucleotide sequence coding for the amino acid sequence of human serum albumin; said nucleotide sequence is as follows:

20

5

25

	(30)	(170)	(192)	(350)	(440)	(330)	(620)	(710)	(300)
	AGG	23 173 AA	20 818 60 60 60 70	8 2 E	110 pro CCA	140 try TAT	170 gln GAA	200 cys TGT	230 g l u CAA
	a E	ohe TTC	phe 111	thr	AAC	leu TTA	169 cys TGC		8 18 CCA
5	3 CTC	AAT	alu I	a1 a	5 S	1 X 7	168 cys TGT	leu lys CTC AAG	phe ala TTT GCA
	phe	n to			GAT	1ys	ale AS		o o o
	3 E	2 YS	CTA	thr val ACA GTT	1ys AAA	lys lys AM AM	ACA	aln ara CAG AGA	ala alu phe GCT GAG TTT
	£	5 5	glu val thr GAA GTA ACT	27 27.5 70.5		T to	phe TTT	lys AA	1ys
10		phe lys asp leu aly TTT AAA GAT TTG GGA	AAT	75 Jeu cys TTA TCC	gin his CAA CAC	phe leu lys lys TTT TTG AAA AAA	ala GCT	ala lys GCC AAA	pro lys ala CCC AAA GCT
		asb GAT		1 ys	leu ain TTG CAA			ser TCT	
		AAA	leu val TTA GTG	SAC CAC	TTC	alu thr GAG ACA	lys ala AAA GCT	36T	arg ACA
		phe		7 53		2 ₹	tyr TAT	318 GCT	220 ser gln arg phe AGC CAG AGA TTT
		10 20 20 20 20 20 20 20 20 20 20 20 20 20	40 val lys GTA AAA	70 phe 111	100 101 glu cys GAA TGC	130 asn qlu AAT GAA	160 arg AGG	190 175 AG	220 3er AGC
15		h1s CAT	his		AAT	GAC	1 ys	919 55	CTG CTG
		ala his GCT CAT	asp GAT	thr leu Acc CTT	25	P.I.s	160 ala lys arg tyr GCT AAA AGG TAT	190 glu gly lys ala GA GCG AG GCT	arg leu CGC CTG
			SA SE		a1 y 666	phe TTT	ohe TT		
		glu val GAG GTT	phe	leu his CTT CAT		ela GCT	phe TTC	arg asp CGC GAT	val ala GTA GCT
20			pro CCA	3e7	glu pro GAA CCT			CTT	
		lys ser AAG AGT	36 Cys 7GT	lys AA	e d	124 cys thr TGC ACT	glu leu leu GAA CTC CTT	골좋	trp ələ TGG GCA
			g1 CAG	339 GAC	1 ys	met ATG	₽. <b>Y</b> 5	asp	-
		ala his GCA CAC	gra	62 cys asp TGT GAC	# 55 SCA	val	6 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 6 7 5 7 5	Jeu CTC	lys ala AAA GCA
		asp GAT	CTT CTT	asn	91 Cys TGT	asp	a) a	lys	Phe TTC
25		- 6 A	30 tyr TAT	8 g 8	8 s 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	120 val GTT	150 tyr TAT	180 57 70	210 glu arg ala phe lys d GAA AGA GCT TTC AAA O
			914 CAG	ala GCT	asp GAC	glu GAG	a ge	leu TG	<u>ئ</u> کې
		phe arg	ala GCT	ser TCA	ala CCT	5 Y	tyr	1eu CTG	glu arg GAA AGA
		p r val GTG	phe	glu GAG			pro CCT	177 cys 100	91.y GGA
		gly	ala	asp	glu met GAA ATG	; arg leu val arg CGA TTG GTG AGA	arg his pro AGA CAT CCT	a1 a	phe TTT
30		arg Acc	11e ATT	ala GCT	919	leu TTG	87.9 AG 43	ala GCT	<u> </u>
		 100 100	1er		tyr TAT	25	arg AGA	133 A	ક ર્
		tyr IAT	leu val leu ile TTG GTG TTG ATT	53 oys val IGT CTT	thr	5 5 5	- <b>*</b> 50	asp lys CAT AAA	ser leu gln lys phe gly AGT CTC CAA AAA TTI GGA
		al a GCT	leu TTG	th ACA	ag 85	leu CTC	ile	al a	ser AGT
35	<b>,</b>	100	21 818 GCC	17.8 AAA	arg CGT	111 asn AAC	141 glu ile GAA ATT	171 ala GCT	201 ala GCC

	(890)	(980)	(1070)	(1160)	(1250)	(1340)	(1430)	(1520)	(1610)	(1700)
5	260 1eu	290 11e ATT	320 a1a GCT	350 ala GCC	380 Leu CCT	410 819 CCT	\$\$0 \$13 CAT	470 ser AGT	SOO 1ys AAA	530 val GTT
	asp GAC	289 cys TGC	tyr TAT	leu CTT	pro CCT	val GTT	₹ ¥	val GTA	0 <u>0</u> 0	
	arg ala AGG GGG	413 CAC	asn	ara AGA	1ys	leu TTA	438 cys TGT	5 Y	val	ala leu GCA CTT
•		3er 707	1ys	leu CTG	alu phe GAA TTT	ala Jeu leu GCG CTG TTA	437 cys TGT			
	asp asp GAT GAC	glu lys	316 val cys GTT TGC	leu Jeu CTG CTG			1ys	AAA	thr tyr ACA TAC	ain thr CAA ACT
10		<b>16</b> ₩	val	leu CTG	asp	asn AAT	ARC	leu his qiu lys thr TTG CAT GAG AAA ACG	a)a GAA	2.3
	ala GCT	pro leu leu CCT CTG TTG	asp	va1 GTG	phe TTC	r te	417	h is	asp (	din ile iys iys CAA ATC AAG AAA
	253 cys TGT	leu CTG	ser lys AGT AAG	val		phe TTC	val GTG	7.5 TG	val GTC	16 /
	of A SA		ser	ser TCT	1ys AAA	1ys	lys val qly AAA GTG GGC	val GTG	glu	ain ile CAA ATC
15	250 253 leu leu glu cys CTG CTT GAA TGT	280 91u 1ys GAA AAA	91u GAA	tyr	ala 000		9 1 y	461 0ys TGT	Jeu g CTG (	
15	250 1eu CTG		310 val GTT	340 asp GAT	370 tyr TAT	400 glu tyr GAG TAC	430 leu qly CTA GGA	460 461 leu oys TTA TGT	490 ala CCT	520 ser glu lys glu arg TCT GAG AAG GAGA
	asp	279 0ys 1GT	phe TTT	pro	369 cys TGC	<del>`</del>		gr	ser i	73 X
	his gly CAT GGA	278 cys TCC	asp GAT	hIs CAT		CTT	arg asn AGA AAC	AAC	phe TTT	leu ser glu 1ys CTT TCT GAG AAG
	n la CAT	16 83	ala GCT	arg AGG	pro hís CCT CAT		3er TCA	Jeu CTG	767	757
20	245 246 cys cys TGC TGC	lys leu lys AAA CTG AAG	ala	arg	pro CCT	glu gln GAG CAG	val	val leu GTC CTG	Pr.o	
	245 cys TGC	Jeu CTG	ser leu TCA TTA	818 GCA	asp CAT	phe TTT	glu val GAG GTC	val GFG	55	514 cys thr leu TGC ACA CTT
	ale GAA		ser TCA	tyr	ala ccA	leu CTT	val	ser val TCC GTG	arg	514 cys TCC
	thr	ser AGT	pro CCT	g]n	ala	ole GAG	leu val CTT GTA		AAC	514 11e cys ATA TGC
	h is	ser TCC	leu TTG	tyr TAT	ala CCC	392 cys TGT	thr	tyr leu IAT CTA	val GTG	asp CAT
25	va1 GTC	270 ser 11e TCG ATC	35 SAC	leu TTG	361 693 TGT	asn	pro	asp CAC	16	818 CCA
	240 173 AAA		300 ala cct	330 phe TTT	360 ays TGC	390 gIn CAA	420 thr ACT	450 910 CAA	480 367	510 h1s CAT
	thr	asp CAT	pro CCT	aet ATG	1ys AAG	1 ys	ser TCA	ala GCA	glu GAA	Phe TTC
	asp leu GAT CTT	gln CAA	glu met GAG ATG	leu gly TTG GGC	leu glu CTA GAG	11e ATC		448 0ys TGT	thr ACA	thr.
30	asp CAT	glu esn GAA AAT				gin asn leu ile CAG AAT ITA ATC	gla	5 CC C	477 cys TGC	phe TTC
	th Ag		asp GAT	val phe GTC TTC	thr	asn AAT	. ord	met ATG	476 eys TGC	AC th
	leu val thr TTA GTG ACA	265 cys TCT	AAT	asp val GAT GTC	thr	gln	val GTA	ACA		glu thr phe GAA ACA TTC
	17A	tyr lle TAT ATC	val glu CTC CAA	asp	28	pro CCT	1ys AA	\$ ₹	thr 1ys ACC AAA	ala GCT
	1ys	tyr TAT	glu val glu GAA GTG GAA	ala lys asp GCA AAG GAT	tyr TAT	glu	1ys AG	ela lys GCA AAA	val	asn (AAT
35	231 val ser lys leu val thr asp leu GTT TCC AAG TTA GTG ACA GAT CTI	lys AAG		ala GCA	th ACA	glu	411 tyr thr lys lys val pro gln val TAC ACC AAG AAA GTA CCC CAA GTG	glu ala lys arg met pro GAA GCA AAA AGA ATG CCC	arg ACA	phe asn
	231 val GTT	261 ala GCC	291 ala GCC	321 91u CAG	351 173 AAG	381 val GTG	tyr TAC	441 pro CCT	471 83P 6	501 glu phe asn GAG TIT AAT

	_	_	
	(1790)	683)	(Z)
	5 8 8 9 8 8 9	570 phe ala glu glu qly lys lys leu val ala ala ser gln ala ala leu qly leu ter TTT GCC GAG GGT AAA AAA CTT GTT GCT GCA AGT CAA GCT GCC TTA GGC TTA TAA CATCACATTTAAAAG (1883)	AGAAAATGAAGGTCAAAAGCTTATTCATCTGTTTTTTTTT
	540  1ys els thr lys glu gln leu lys els val met asp asp ohe als als phe val glu lys cys cys lys AAG GCA ACA AAA GAG CAA CTG AAA GCT GTT ATG GAT CAT TTC GCT GTT TTT GTA GAG AAG TGC TGT AAG (	er AAA	ATT
5	8 8 5 7 € 5 7	ATTT	777
	× × × × × × × × × × × × × × × × × × ×	TCAC	TAAA
	2 A	5	MCA
	ร์ธ์	7 te	WAA.
	3 5	1 1 1	701
10	ž E	6 9 1	000
	₹ <u>5</u>	3 5	ACAC
	# 55 # 55	<b>8</b> 200 .	<b>5</b> 22
	1 6	8 13 CCT	TAA
15	550 asp CAT	580 g1n CAA	ວະວ
13	\$ 5 5	ser AGT	ret <sub>T</sub>
	Met ATG	ور درک	Ħ
	c t	al a	77.
	SCT SC	val GTT	3111
20	\$ <del>2</del>	S T C	VTCT
	leu CTG	lys AAA	110
	ole CAA	1 × ×	CTTA
	916	aly GGT	. AAG
	.¥ 8 <b>₹</b>	91u GAG	ATCA
25	S&O thr	570 glu GAG	8
	# to	13 200	MAT
	1ys AG	phe TT	¥ĕ
	 	567 cys TGC	<u>ა</u>
20	1ys AAG	thr	ter ter TCACAATAA
30	ς <del>Ι</del> .	glu GAG	20
	1 × s	lys MG	CAT.
	leu val lys his lys pro CTC GTG AAA CAC AAG CCC	ssp asp lys glu thr cys GAC GAT AAG GAG ACC TGC	CTAC
	10 V	S b c	נאפנ
35	531 91u 1 GAG C	561 ala a CCT C	ter ter Catctcagcctaccatgagataagagaa
J	G G	v 10 C	Ü

TCATTITGCCTCTTTTCTCTGTGCTTCAATTAATAAAAATGGAAAGAATCTAA.... 20 ....AA (2078)

6. Nucleotide sequence of the cDNA of human serum albumin, said nucleotide sequence is as follows:

5	20 1ys AAA (170)	(260)	(350)	(440)	(330)	(620)	(710)	(300)
		20 20 31 31 32 32 32 32 32 32 32 32 32 32 32 32 32	80 CTT	110 pro CCA	140 try TAT	170 GAA	200 cys TGT	230 910 GAA (
	phe TTC	phe	thr	asn AAC	leu TTA	169 cys TGC	lys AG	8 1 8 CC 7
	asn	ale CAA	818 GCA	aso GAC	TAC	168 cys TGT		phe TTT
10	glu asn GAA AAT	thr	va] GTT	asp GAT	lys tyr AAA TAC	ole GAA	ara leu AGA CTC	alu c
10	£ 85	val	thr	13s	3 × 8	thr ACA		ala CCT
	5 5	75. 649	25 25 100	h is CAC	leu lys TTG AAA	phe TT	lys aln AAA CAG	lys ala AAA GCT
	1 tc	asn	leu TTA	cy di			818 CCC	570
	asp	val	lys leu AAA TTA	leu aln TTG CAA	thr phe ACA TTT	ala ala GCT GCT	36T	phe pro TTT CCC
15		TY T	asp GAC	phe TTC		AA	3er 700	
	phe lys TTT AAA	lys AAA	91y CGA		alu alu CAA CAG	tyr	ala	gin arg CAG AGA
	10 his arg CAT CGG	40 CTA	2 phe 11 1	S & 50	130 asn AAT	160 arg AGG	190 1ys AAG	
		S F S	le. CTT	AAT	asp	1ys AA	2 5 5 5 5 5 5 5	220 Jeu ser CTG AGC
	val ala GTT GCT	asp CAT	thr	87 AGA	h s	ala GCT	2€	
20		a ja	h 13	9) 900	phe TT	ohe TTT	asp	ala arg GCT CGC
	lys ser glu AAG AGT GAG	T T	ser leu TCA CTT	pro	ala GCT	leu phe CTT TTC	<b>31.</b> 9	val
	ser	Pro		g Ju GAA	124 cys thr TGC ACT		leu	ala GCA
		34 gin eys CAG IGI	<u>₹</u>	gla	124 cys TGC	glu leu GAA CTC	£ 8	
25	ala his GCA CAC	gln CAG	asp	Jys AA	met ATG	glu GAA	asp	ala trp GCA TGG
	1 asp ala GAT GCA	gln CAG	62 973 TGT	هاه درک	asp val GAT GTG	<b>Pr</b> 0	leu CTC	1ys
	asp CAT	30 tyr leu TAT CTT	AAT	91 035 161	asp	818 CCC	73 AG	phe
		30 tyr TAT	glu SA	90 cys 160	120 glu val GAG GTT	150 tyr TAT	180 pro	210 818 GCT
		gln CAG	ala CCT	asp GAC	ole GAG	phe	leu TTG	glu arg
30		ala GCT	ser TCA	ala GCT	pro CCA	tyr TAC	leu CTG	ale GAA
		phe	glu	glu met CAA ATG	Ag arg	his pro CAT CCT	177 Iys ale ale oys AAA GCT GCC TGC	gly GGA
		ala ccc	asp GAT	ag 8	arg leu val arg CGA TTG GTG AGA	his CAT	a 18	phe TTT
		11e ATT	818 GCT	gly	arg leu CGA TTG	arg AGA	ala GCT	17.8
35		leu TTG	vai CTT	tyr	5 S	arg AGA	15 A	gla CAA
33		leu val leu lle TTG GTG TTG ATT	53 595 767	thr	pro CCC	fle ala arg arg ATT GCC AGA AGA	asp GAT	leu gin lys phe gly CTC CAA AAA TTT GGA
		leu TTG	th ACA	9 to	leu CTC		ala GCT	ser
		21 a1a GCC	2 <sup>1</sup> √3 ×3	81 arg	111 asn AAC	14.1 SA	171 ala CCT	201 a1a GCC

5	(1890)	(980)	(1070)	(1160)	(1250)	(1340)	(1430)	(1520)	5n0 1ys AAA (1610)	(1700)
	260 leu CTT	290 11e ATT	320 ala CCT	350 ala GCC	380 leu CCT	410 ara	440 h1s	\$70 \$er .	500 1ys AAA (	
	asp GAC	289 cys 100	tyr TAT		5 5	val GTT	1 ys 1		pro 1	530 leu val CTT GTT
	818 606	h is CAC	asn	2 2	1ys AAA	leu v	438 cys lys TCT AAA	pro val CCA GTA		7 P. C.
10	£ 55 ₹	ser Tor	1ys AAA	leu CTG	phe lys TTT AAA	Jeu	437 cys 161	thr p	tyr val TAC GTT	thr ala ACT GCA
	asp asp arg ala asp GAT GAC AGG GCG GAC	glu lys ser his GAA AAA TCT CAC	316 cys TGC	re CTG	o Jo CAA	818 GC	1ys		thr t	₹ ¥
	ase GAT	ale GAA	va] GTT	ser val val leu leu leu ara leu TCT GTC GTG CTG CTG AGA CTT	330	AAT	ser lys AGC AAA	alu 1ys GAG AAA	alu thr Gaa aca	530 gin fle lys lys gin thr ala leu val CAA ATC AAG AAA CAA ACT GCA CTT GTT
	ala CCT	pro leu leu c CCT CTG TTG	ser lys asp val AGT AAG GAT GTT	val GTG	ake 110	u e Cyc			asp q	ys - AG A
	253 glu eys GAA TGT	pro leu CCT CTG	lys AAG	va] GTC	val GTG	phe TTC	val aly GTG GGC	leu his TTG CAF	val a	16 J
15	250 leu leu glu CTG CTT GAA				1ys	1ys	1ys	val 1 GTG 1	glu val GAA GTC	1 4 X
	leu CTT	280 qlu lys GAA AAA	310 val glu CTT GAA	tyr TAC	a] a ဝင်င		45	\$61 cys v		ខ្
		280 q1u C&A	310 val CTT	340 asp CAT	370 tyr TAT	400 glu tyr GAG TAC	430 ser arg asn leu qly lys val TCA AGA AAC CTA GGA AAA GTG	460 461 leu cys val leu his TTA TGT GTG TTG CAT	ala leu GCT CTG	520 glu arq GAG AGA
	gly asp GGA GAT	279 cys TGT	phe	pro	369 cys TGC	<del>2</del> 3	asn AAC			
20		278 cys TGC	asp CAT	ele arg arg his GCA AGA AGG CAT	a Jr CAA	leu CTT	ACA ACA	leu asn gln CTG AAC CAG	pro cys phe ser CCA TGC TTT TCA	leu ser glu lys CTT TCT GAG AAG
	245 246 glu cys cys his GAA TGC TGC CAT	leu lys glu CTG AAG GAA	leu ala ala TTA GCT GCT	arg AGG	his CAT	al CAG	ser TCA	1 ec	160	7 5 T
	246 cys TGC	leu lys CTG AAG	ala GCT	arg AGA	pro CCT	alu GAG		val GTC	679 CCA	3 = =
	245 Cys TGC	leu CTG			asp CAT	phe TTT	glu val GAG GTC	ser val val TCC GTG GTC	5 25	
	£ 8	ser lys AGT AAA	pro ser CCT TCA	glu tyr GAA TAT	ala GCA	leu CTT		3er TCC	arg Acc	514 cys thr TGC ACA
25	his thr CAC ACG	ser AGT	pro CCT		ala GCT	g to	leu		asn arg AAC AGG	11e c
		ser TCC	leu TTG	tyr	13 CC	392 cys TGT	thr	tyr leu TAT CTA		esp CAT
	240 1ys val AAA GTC	11e	asp GAC	1ec 776	361 cys TGT	390 gln esn CAA AAT	Pro CCA		680 ser leu val TCC TTG GTG	
		270 3er 1CG	300 818 CCT	330 phe TTT	360 cys 760		420 thr ACT	450 glu asp GAA GAC	\$80 7CC	510 his ala CAT GCA
20	ACC Th	83p CAT	pro	met ATG	1ys AG	1ys	ser TCA	a1 a	glu	phe I
30	asp leu GAT CTT	gla	glu asn asp glu met GAA AAT GAT GAG ATG	91 <i>y</i>	leu glu CTA GAG	11e ATC	val GTG	448 033 1GT		
	38	AAT	28	leu TTG	glu thr thr leu GAA ACC ACT CTA	gin asn leu ile CAG AAT TTA ATC	pro gln val	979 CCC	677 cys TGC	glu thr phe thr GAA ACA TTC ACC
	thr ACA	916	asp CAT	phe TTC	thr	asn	010	me t ATG	476 477 cys cys TGC TGC	thr I
	ser lys leu val TCC AAG TTA GTG	265 11e cys ATC TGT	asn AAT	val GTC	thr	gln CAG	vai GTÅ	ala lys arg GCA AAA AGA	1,3 <b>A</b>	914
35	3 E			asp CAT	916	gtu pro GAG CCT	1. XX	173 AA	ACC	ala GCT
	1 1 ys	lys tyr AAG TAT	glu val GAA GTG	1ys AAG	tyr IAT	g to	1ys AAG	ala lys GCA AAA	val	asn AAT
	3er			321 glu ala iys asp val phe leu gly GAG GCA AAG GAT GTC TTG GGC	thr	91c 6A	411 tyr thr TAC ACC	910	476 477 arg val thr lys cys cys thr AGA GTC ACC AAA TGC TGC ACA	phe a
	231 val GTT	261 ala GCC	291 ala GCC	321 91u CAG	351 1ys AAG	381 val GTG	411 tyr 1AC	441 pro	471 889 GAC	501 glu phe asn ala GAG TTT AAT GCT

5	1 559 560 1 0ys 1ys 1 TGC AAG (1790)	F AAAG (1883)	TTAA (2002)
10	558 val glu lys cys GTA GAG AAG TCC	580 ser gin ala ala leu qiy leu ter AGT CAA GCT GCC TTA TAA CATCACATTTAAAAG (1883)	ter ter ter Cateres de Ler Cateres de Les de
15	550 lys glu gin leu lys ala val met asp asp phe ala ala phe v AAA GAG CAA CTG AAA GCT GTT ATG GAT GAT TTC GCT GCT TTT (	580 r gin ala ala leu q' r CAA GCT GCC TTA G	TGGTGTAAAGCCAACACCC
20	lys ala val met as  AAA GCT GTT ATG GA'	570 ala glu glu qly lys lys leu val ala ala ser gln ala ala GCC GAG GAG GGT AAA AAA CTT GTT GCT GCA AGT CAA GCT GCC	4161611117611117661
25	0 r lys glu gln leu A AAA GAG CAA CTG	0 u glu qly lys lys .G GAG GGT AAA AAA	AGATCAAAAGCTTATTC/
30	540 iys pro lys ala thr 1 NG CCC ANG GCA ACA A	phe TTT	ter ATAAGAGAAAGAAATGA
35	531 glu leu val lys his lys pro GAG CTC GTG AAA CAC AAG CCC	561 ala asp asp lys glu thr cys GCT GAC GAT AAG GAG ACC TGC	ter ter CATCTCAGCCTACCATGAGAATAA

TCATTITGCCTCTTTTCTCTGTGCTTCAATTAAAAATGGAAAGAATCTAA.... 20 ....AA (2078)

7. Nucleotide sequence coding for the prepeptide of human serum albumin, said nucleotide sequence is as follows:

	a i buin i i i ,	Said	nucleotige	sequence	15
				(30)	
5				AGC.	
				-10 lle ser leu leu phe leu ohe ser ATT TCC CTT CTT TTT CTC TTT AGC	
				1 5 C	
				phe TTT	
10				CT CT	
				-10 1 mg/	
				36. 707	
			•	116	
				ag E	
15				p r o trp val tlu phe 1 TGG GTA ACC TTT A	
				3 % a 23 a	
				5 tr	
				18 16 13	
20				۱ <del>۲</del> ۷	
		-		ZGCA	
				CCTTI	
				د٧ود	
25	•			<b>Y</b> 233	
				.caac	
				-18 Met lys FTCTCTCTGTCAACCCCACAGCCTTTGGCACA ATG AAG	7
				CTCT	
				667777	0
30				8	۵
					9
					-
					•
35					

35

ser ale tyr ser arg gly val phe arg arg

8. Nucleotide sequence coding for pro human serum albumin, said nucleotide sequence is as follows:

5	(170)	(260)	(350)	(440)	(330)	(620)	(710)	(300)
	20 1ys	So ala GCA	80 Jeu CTT	110 Bro	140 try TAT	170 q1n CAA	200 cys TGT	230 910 GAA (
	phe TTC	Dhe TT	thr	AAC	1eu	169 Cys TG:		מנא פ
	asn	alu CAA	818 GCA	aso GAC	tyr TAC	168 cys TGT	leu lys CTC AAG	phe
	ole GAA	thr	val	GAT		age 45		g to
10	9 tu	crA	thr	1ys	1ys	thr	aln ara CAG AGA	ala GCT (
		o Jo		CAC	phe leu lys lys TTT TTG AAA AAA	phe TT		AAA
	leu aly TTG GGA	AAT	Jeu TTA	gln CAA	phe TTT	ala phe GCT TTT	ala lys GCC AAA	Pro CCC
	asp		1 ys	16 176		ala ccr	36T 7	ahe TT
15	17s	leu val TTA GTG	75 asp lys leu oys GAC AAA TTA TGC	Dhe TTC	glu thr GAG ACA	1ys	3 er	ACA .
	phe lys asp TTT AAA GAT		91y CGA	101 105 100	glu GAA			gln
		40 val 1ys GTA AAA	70 phe TTT	5 a 3 8	130 asn AAT	160 arg tyr AGG TAT	190 lys ele AAG GCT	220 3er AGC (
	10 hIs arg CAT CGG	h is		AAT	asp GAC		91° 700 800	leu CTG
	ala GCT	CAT	thr leu ACC CTT	2	hts	ala lys GCT AAA	₹ <b>3</b>	arg leu CGC CTG
20	val CTT	SAA S		4 p	phe TT	ohe TTT	GAT	ala GCT
	lys ser glu val ala his AAG AGT GAG GTT GCT CAT	phe	lys ser leu his AAA TCA CTT CAT	pro CCT	318 GCT	phe TTC	879 CG	val
	aer AGT	pro	ser TCA	976	thr	leu phë CTT TTC		ala GCA
	1ys AAG	34 cys TGT	lys ser AAA TCA	gin glu CAA GAA	124 cys TGC	pro glu leu leu phe CCG GAA CTC CTT TTC	glu leu GAA CTT	trp ala TGG GCA
25		gln cAG	asp		ATG	36,85 CA	asp	ala GCA
25	818 his GCA CAC	gla	62 oys asp TGT CAC	ala lys GCA AAA	val	2 9 2 9	5 CTC	1ys AAA
	asp GAT	1eu CTT	asn	91 Cys	asp	# D	1ys AAG	phe TTC
	- £ 50	30 tyr TAT	60 916 6A	90 93 75C	120 val GTT	150 tyr TAT	180 pro	210 ala GCT
	o phe arg	gla			g r GAG		1 cu	arg AGA
30	o ag I	ala GCT	ser ala TCA GCT	ala asp CCT GAC	5 Y	tyr phe TAC TTT	)eu CTG	gyv Cyv
	ריי מזמ	phe	g pro		arg AGA	Pro CCT		917
	p r gly val GGT GTG	ala ၁၁၁	asp	glu met GAA ATG	val	h13 CAT	a1 a	phe TTT
	91.6 ACC	11e ATT	ala CCT	gly CCT	leu val TTG GTG	arg hís AGA CAT	ala GCT	1ys
		leu TTG		tyr	<del>د</del> و ع	4 5 A	\$ ₹	gln
35		val leu lle GTG TTG ATT	53 cys val TGT GTT	thr	5 2	318 CCC	asp	leu gin lys phe CTC CAA AAA TTT
		1eu 77G	thr	of <b>X</b>	leu CTC	ile ATT	ala GCT	AGT
		21 818 GCC	12 17 A	81 arg CCT	111 asn AAC	141 914 6A	171 ala GCT	201 ala :

5	(890)	(980)	(1070)	(1160)	(1250)	(1340)	(1430)	(1520)	(1610)	(1700)
	260 1eu CTT	290 11e ATT	320 ala GCT	350 ala GCC	380 leu CCT	\$10 810 CGT	860 813 CAT (	470 36F AGT (	Sno 17s AAA C	_
	ala asp OCG CAC	289 cys TGC	tyr		Bro CCT		lys I	_	Pro 1	530 leu val CTT GTT
		M IS	asn tyr AAC TAT	ř. 73	1ys	TTA T	438 cys TCT	pro val CCA GTA	val p	ala leu GCA CTT
10	asb arg GAC AGG	280 glu lys pro leu leu glu lys ser GAA AAA CCT CTG TTG GAA AAA TCC	316 glu ser lys asp val cys lys GAA AGT AAG GAT GTT TGC AAA	leu leu leu arg leu CTG CTG CTG AGA CTT	phe TTT	leu leu val CTG TTA GTT	437 cys	thr p	490 ala leu glu val asp glu thr tyr val pro GCT CTG GAA GTC GAT ACA TAC GTT CCC	hr CT G
	255 CAC	1ys AAA	316 cys TGC	<u> </u>	asp glu phe GAT GAA TTT	2 a l a	lys c	× × ×	thr t	ain ile lys lys ain the CAA ATC AAG AAA CAA ACT
	\$30 GAT	alu CAA	val	leu leu CTG CTG	asp CAT	lys phe gin asn ala AAA TTC CAG AAT GCG	ser lys AGC AAA	alu lys GAG AAA	alu t	% <b>₹</b>
	ala SCT	leu TTG	asp GAT	val		ata SAG	7 2 33 33 3		SSD 0	5 S
	253 glu cys GAA TGT	Jeu CTG	1ys AAG	val GTC	val phe GTG TTC	phe c	val qly GTG GGC	leu hís TTG CAT	val a GTC G	. in 17.
15		pro CCT	ser AGT	ser TCT		£ <del>X</del>	lys AAA (		91c v	- X
	250 leu leu CTG CTT	1ys AAA	916 684	7 t	ala lys GCC AAA	tyr TAC /	aly lys val aly GGA AAA GTG GGC	461 cys val TGT GTG	leu g CTG G	5 S
	250 1ed CTG		310 val CTT	340 38p GAT	370 tyr TAT	400 glu tyr lys GAG TAC AAA	436 lev c CTA (	460 4 Jeu c TTA T	490 ala 1 GCT C	520 glu arg gln lle lys GAG AGA CAA ATC AAG
	SAT CAT	279 cys TCT	phe TTT	pro CCT	369 cys TGC	<del>2</del> 29	436 ser arg asn leu TCA AGA AAC CTA	gln 1 CAG T		χς Ας α ν
20	his gly CAT GCA	278 978 160	asp GAT	M s CAT		leu CTT	arg .	asn g	phe ser TTT TCA	glu lys GAG AAG
20		91°	ala GCT	arg hís AGG CAT	his CAT	2 CAS	3er TCA	leu esn CTG AAC		er g
	245 246 glu cys cys GAA TGC TGC	278 Ser lys leu lys glu oys AGT AAA CTG AAG GAA TGC	leu ala TTA CCT	ala arg GCA AGA	pro	a)u GAG		val	pro cys CCA TGC	514 asp ile cys thr leu ser glu lys CAT ATA ICC ACA CTT TCT CAG AAG
	245 glu cys GAA TGC	CTC			asp GAT	phe TT	glu val GAG GTC	val v GTG 0		- 5 - 5
		1ys AA	pro ser CCT TCA	glu tyr GAA TAT	ala CCA	leu CTT	val GTA	3er 700	asn arg arg AAC AGG CGA	514 cys t TCC A
25	thr Acc	AGT	pro CCT	glu	ala GCT	g r GVe	thr leu val ACT CTT GTA		asn a	7 5 T
	N 15	<b>16.</b>	leu TTG	leu tyr TTG TAT	a1a ೧୯୯	392 cys TGT	thr ACT	tyr leu TAT CTA	val a	asp lle CAT ATA
	240 lys val AAA GTC	11e ATC	300 ala asp leu GCT GAC TTG	leu 77G	361 cys TGT	AAT		esp CAC	leu v TTG (	ala a CCA C
•	240 178 AA	270 ser TCG		330 phe 111	360 373 155	330 g1n CA	420 thr pro ACT CCA	450 91u CAA	480 ser leu val TCC TTG GTG	510 hís a CAT G
	thr Acc	asp CAT	pro CCT	met ATG	1ys AAG			25		he h
30	15 CT	glu asn gln GA AAT CAA	glu met CAC ATC	phe leu gly met p TTC TTG GGC ATG 1	tyr glu thr thr leu glu lys TAT CAA ACC ACT CTA GAG AAG	asn leu 11e lys AAT TTA ATC AM	gin val ser CAA GTG TCA	648 633 TCT	476 477 thr lys cys cys thr glu ACC AAA TGC TGC ACA GAA	phe thr phe I
	thr asp ACA GAT	88n AAT	glu Ge	Jeu TTG	Jeu CTA	asn leu ile AAT TTA ATC	pro gln vel CCC CAA GTG	5 5	477 cys t TGC /	he t TC A
	thr ACA	ole CA	asp GAT	phe TTC	thr	asn	. 67 <b>9</b>	net	476 cys 7CC	thr p
	ser lys leu val TCC AAG TTA GTG	265 ile ays ATC TGT	asn AAT	ala lys asp val GCA AAG GAT GTC	thr Acc	g], CAG		arg met AGA ATG	\$ <del>\$</del> \$	glu thr GAA ACA
35	1eu TTA	11e ATC	glu GAA	asp GAT	88	glu pro CAG CCT	lys lys vel AG AA GTA	/ys /	hr 1	ala g CCT G
	1 ys	tyr	glu val GAA GTG	lys AG	tyr IAI	glu	lys MG	ele lys GCA AAA	val t	asn ala AAT GCT
		1ys				onle SA	thr 1	glu ala lys arg met pro GAA GCA AAA AGA ATG CCC	arg v	ie a
	231 val GTT	261 ala GCC	291 #14 GCC	321 glu GAG	351 1ys thr AAG ACA	381 val GTG	411 tyr thr TAC ACC	pro g	471 asp arg val GAC AGA GTC	glu phe

	6	6	2
5	(1790)	(188	(200
		¥€	14
	559 560 cys lys TGC AAG	ter TTAA/	TCTT
	558 0ys 1GC	ACAT	AATT
	lys AAG	CATC	CATA
10	3AG	TAA	*
	781 STA (	leu ter TTA TAA CATCACATTTAAAAG (1883)	CTAA
	ag E	× 88	CTGT
	ala p	TA C	CACC
15	12 2		CCAA
15	3 Pe	ala CCT	AAAG
	558 . glu gin ieu lys ala val met asp asp phe ala ala phe val glu lys oys ogs CAG CAA CTG AAA GCT GTT ATG GAT GAT TTC GCT GCT TTT GTA GAG AAG TGC	570 ala glu glu qly lys lys leu val ala ala ser gln ala ala leu qly leu ter GCC GAG GAG AAA AAA CTT GTT GCT GCA AGT CAA GCT GCC TTA GAC TAA TAA	CTGT.
	GAT	AGT	СТТС 2078
	ATG	ala GCA	TTT0
20	val GTT	ala CCT	1011
	als GCT	val	20 .
	1.75 AA	Jeu CTT	VTCTC
	1eu CTG	1ys AAA	VTTC/
0.5	g t	<b>₹</b> ₹	XTT/
25	o se co	aly GGT	AAG,
	1ys	g)u GAG	ATC/
	540 818 thr 1y8 g GCA ACA AAA G	570 91u GAG	IGAAC VAAA
	ala GCA	a1a GCC	MAAI
30	1ys AAG	phe TTT	WAG/
	pro CCC	567 cys TGC	AGAG
	17s	thr	tei MTA/ TGTG
	531 glu leu val lys his lys pro lys GAG CTC GTG AAA CAC AAG CCC AAG	567 ala asp asp lys glu thr cys phe GCT GAC GAT AAG GAG ACC TGC TTT	ter ter Catctcacctaccatgagaataagaaaatgaagatcaaagcttattcatctgtttttttt
	1ys AAA	1ys AAG	ACCA)
35	val GTG	asp GAT	GCCT.
	leu CTC	asp GAC	CTCAI
	531 910 GAG	561 ala GCT	CAT

9. Nucleotide sequence coding for the pre pro human serum albumin, said nucleotide sequence is as follows:

5	(30)	(170)	(1260)	(350)	(440)	(330)	(420)	(710)	(300)
	AGC	20 1ys	SO ata	5 T T	110 Bro	140 try TAT	170 91n CAA (	200 cys TGT (	230 91u GAA (
	a E	phe TTC	phe TTT	thr	AAC	1 c	169 cys 1GC	lys o	هاه در م
	150	AAT	ale CAA	£ 200	38.0	177	168 cys TGT	lea 1 CTC /	
10	phe TTT	D AA	thr	val GTT	asb CAT	k A	alu CAA	2	alu phe GAG TTT
10		a fo		th ACA		lys AAA		4 6 7	ala GCT C
	-10 lev lev CTT CTT	ور ورد ع	alu val CAA GTA	55 ST	his lys CAC AAA	leu lys TTG AAA	phe thr TTT ACA	lys ain ara AAA CAG AGA	lys ala AAA GCT
	<b>2</b> 20 20 20 20 20 20 20 20 20 20 20 20 20	1 tc	AAT	leu TA	£ 5				
	11e	asp GAT	val GTG	1ys	Jeu 11G	thr phe ACA TTT	ala ala GCT GCT	ser ala TCT GCC	phe pro TTT CCC
15	a be	¥.¥	3 X	asp CAC	ohe TTC	glu glu thr phe GAA GAG ACA TTT	XX XX		ACA 1
	p r c trp val tlu phe TGG GTA ACC TTT	phe	1 ys	÷ 3	5 % 5 %	£ 8	tyr TAT	190 lys ala ser AAG GCT TCG	220 leu ser gin arg CTG AGC CAG AGA
,	CT A TO	5 878 883	40 Val GTA	20 Phe	100 101 glu cys GAA TGC	130 asn AAT	160 879 AGC	130 AG	220 Ser
		MIS CAT	h is CAT	leu	asn	asp GAC	<u>}</u>		220 leu ser CTG AGC
00	1ye	els CCT	asp GAT	th Age	<u> </u>		ala CCT		<b>9.1</b> 0
20	-18 ATG	val GTT	6 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8	his CAT	919 000	phe	ohe	asp glu GAT GAA	ala arg ser cse
		gju	phe TTT	2 5	<b>010</b> CCT	ala GCT	phe TTC		
		AGT	Pro CCA	ser 1CA	es es	thr	leu CTT	glu leu arg GAA CTT CGG	ala trp ala val CCA TGG GCA GTA
		ly.	34 Cys	lys AA	gl,	124 cys TGC	leu CTC		trp 166
25		418 CAC	gln	asp	₹ <b>X</b>	met ATG	ag &	asp	ala trp GCA TGG
		<b>?</b> 5	91,	62 cys TGT	<b>₽</b> 55	val	ala pro OCC CCG		1ys
		GAT GAT	Je CTT	asn	91 cys TGT	asp GAT	# D	180 pro lys leu CCA AAG CTC	210 ala phe GCT TTC
		- 🗜 છુ	30 tyr TAT	60 15 54	8 8 55 T	120 val GTT	150 tyr TAT		210 ala GCT
		pro glyval phe arg GGT GTG TTT CGT	56	ala GCT	asb GAC	o Je CYC	phe TTT	177 1ys ala ala cys leu leu AAA GCT GCC TGC CTG TTG	210 arg ala AGA GCT
30		. 4g ::	ala GCT	ser TCA	ala ccr	£ 55	tyr TAC	of C	ale GA
		7 € 1 GTG	phe TTT	glu	glu met GAA ATG	2. A.	pro CCT	177 eys TGC	91y GGA
		91 <i>y</i> GGT	هاه 90	asp GAT	g &	val GTG	h1s CAT	177 ala cys GCC TGC	phe TTT
		9- 9-	11e ATT	ala GCT	913	arg leu val arg CGA TTG GTG AGA	arg arg his pro AGA AGA CAT CCT	ala GCT	1ys
35		- <b>-</b> 55	1ev 11G	53 cys val IGI GIT	tyr TAT	5 2	\$ 5 V	1ys	투 장
55		tyr	<b>₹81</b>		thr Acc	5 C	ala OCC	esp GAT	leu gin lys phe gly glu CTC CAA AAA TTT GGA GAA
		•1• CCT	]ec	thr	91. CA	leu CTC	ile ATT	ala GCT	Ser
		<b>367</b>	21 ala CCC	22 173 8	81 819 CCT	111 asn AAC	14.1 910 GAA	171 818 801	201 ala GCC

5	(890)	(980)	(1070)	(1160)	(1250)	(1340)	(1430)	(1520)	(1610)	(1700)
	260 1eu CTT	290 11e ATT	320 ala CC.T	350 818 CCC	380 leu CCT	410 ara CGT	A40 his cat	470 ser AGT	570 175	530 val GTT
10	asp GAC	289 cys TGC	tyr TAT	lea CTT	DTO CCT	val GTT	1ys AAA	val GTA	5 TO CO	leu CTT
	a la 000	h1s CAC	AAC	arg AGA	1ys AAA	leu val TTA GTT	438 cys TGT	010 CCA	tyr val TAC GTT	gin thr ala CAA ACT GCA
		ser TCC	1ys AA	leu CTG	phe TTT	ala leu GCG CTG	437 cys TGT	Acc	ty 7⊀7	thr
	asp arg	1ys AAA	316 asp val cys GAT GTT IGC	leu CTG	alu GAA		lys AAA	1ys .	alu thr GAA ACA	£ S
	ala asp CCT CAT	leu alu TTG GAA	va! GTT	Jeu CTG	asp	AAT	430 leu qly lys val qly ser CTA GGA AAA GTG GGC AGC	460 461 leu cys val leu his glu TTA TGT GTG TTG CAT GAG	gle GAA	1ys AAA
		Jeu 116		val	phe TTC	ele CAG	4 86 86	h1s CAT	asp	1ys AAG
	253 glu cys GAA TGT	pro leu CCT CTG	ser lys AGT AAG	val GTC	lys val	lys phe AAA TTC	val	leu TTG	glu val GAA GTC	ain ile CAA ATC
15	ag GA	878 CCT	Ser		1 ys		1ys AAA	val GTG		£ \$
	250 leu leu CTG CTT	280 qlu lys GAA AAA	310 val glu GTT CAA	340 asp tyr GAT TAC	၁၂၈ ၁၁၁	400 glu tyr GAG TAC	a) GGA	461 cys TGT	490 ala leu CCT CTG	520 glu arq GAG AGA
			310 val CTT		370 tyr TAT			\$60 1eu 11A		520 91u GAG
	GAT	279 cys 1GT	asp phe GAT TTT	pro CCT	369 qlu cys GAA TGC	leu aly CTT GGA	arg asn AGA AAC	asn gln AAC CAG	phe ser TTT TCA	glu lys GAG AAG
20	91. 66.	278 cys TCC		his CAT		leu CTT	8 TG ACA	AAC		
20	245 246 glu oys oys his GAA TGC TGC CAT	lys leu lys glu AAA CTG AAG GAA	ata CCT	arg AGG	h is	419	ser TCA	Jeu	pro cys CCA TGC	leu ser CTT TCT
	246 973 150	lys AG	ala CCT	arg AGA	pro CCT	GAG GAG	glu val GAG GTC	val		
	245 glu oya GAA TGC	Jeu CTG	Jeu 17A	ala GCA	asp.	phe TTT		va1 GTG	55	514 ays thr TGC ACA
			ser TCA	tyr TAT	818 GCA	leu CTT	leu val CTT GTA	36r 100	arg	514 678 160
25	his thr	ser	pro CCT	glu GAA	ala CCT	ole GAG	leu CTT	1eu CTA	asn	11e ATA
		ser TCC	Jeu 116	tyr IAT	ala GCC	392 cys 1GT	thr	tyr TAT	leu val TTG GTG	asp GAT
	240 1ys vel AAA GTC	11e ATC	43 CA CA	1eu 11G	360 361 0ys cys TGC TGT	asn AAT	pro CCA	GAC GAC	leu 17G	ale CCA
		270 ser 100	300 ala GCT	330 phe TTT		390 91n CAA	420 thr	450 g1u GAA	480 ser	510 his
	thr	asp	pro CCT	met ATG	1ys AAG	1. ₹	Ser	ala GCA	glu	phe TTC
30	3 F	asn gln AAT CAA	glu met GAG ATG	91y GGC	leu glu CTA GAG	leu ile TTA ATC	gin val CAA GTG	448 cys TGT	thr	AC th
	\$ 5 \$ 7		91e GAG	leu 17G		gin asn leu ile CAG AAT TTA ATC	gln	979 555	477 6ys 1GC	glu thr phe thr GAA ACA TTC ACC
	th VCA	265 cys glu TGT GAA	asp	phe TTC	thr ACT	gin asn CAG AAT	val pro	met ATG	476 1ys.cys AAA TGC	thr ACA
35	<b>å1</b> GTG		AT	val GTC	thr	ole CAG	val GTA	arg AGA	1. ₹	g fu GAA
	ser lys leu val thr asp leu TCC AAG TTA GTG ACA GAT CTT	tyr lle TAT ATC	glu val glu GAA GTG GAA	ala lys asp val phe leu gly GCA AAG GAT GTC TTC TTG GGC	tyr glu thr thr TAT GAA ACC ACT	glu pro GAG CCT	lys lys val Aag aaa gta	glu ala lys arg met GAA GCA AAA AGA ATG	476 477 arg val thr lys.cys cys thr AGA GTC ACC AAA TGC TGC ACA	phe asn ala glu thr phe thr TTT AAT GCT GAA ACA TTC ACC
	1,4 AAG	tyr TAT	val	1ys AG	tyr 1AT	glu GAG	lys AG	ala GCA	val GTC	AAT
		lys AAG		ala CCA	351 1ys thr AAG ACA	381 val glu glu pro GTG GAA GAG CCT	thr		arg AGA	501 glu phe GAG TTT
	231 val	261 ala GCC	291 ala ccc	321 glu GAG	351 1ys AAG	381 val GTG	411 tyr TAC	##1 pro	471 33P GAC	501 g1u GAG

TCATTTTGCCTCTTTTCTCTGTGCTTCAATTAATAAAAATGGAAAGAATCTAA.... 20 .....AA (2078)

5	(1790)	ala glu glu gly lys lys leu val ala ala ser gln ala ala leu gly leu ter GCC GAG GAG GGT AAA AAA CTT GTT GCT GCA AGT CAA GCT GCC TTA GGC TTA TAA CATCACATTTAAAAG (1883)	(70)7
	560 1 Jys	<u>و</u> ۲	5
	χ <del>ξ</del> δ	ter TTAA	=
	558 cys c	ACAT	-
10	Iys	CA16	
	glu Iys	ter TA	כ כ כ
	ala ala phe val	leu TTA	
	phe TTT	41y GGC	5
	ala GCT	15 77 75 75 75 75 75 75 75 75 75 75 75 75	֝֝֝֓֓֓֓֓֓֓֓֓֓֓֓֓֓֓֓֡֓֓֓֓֓֡֓֓֡֓֡֓֡֓֓֓֡֓֡֓֡֓֡
15	818 GCT	300	ב ב
	550 asp phe	580 gin ala ala leu qiy leu ter CAA GCT GCC TTA GGC TTA TAA	{
	550 cAT	580 91n CAA	3
	met asp ATG GAT	910 glu gly lys lys leu val ala ala ser o GAG GAG GGT AAA AAA CTT GTT GCT GCA AGT o	5
20	540 ala thr lys glu gln leu lys ala val met GCA ACA AAA GAG CAA CTG AAA GCT GTT ATG	ala GCA	
20	8 val	. GCT	;
		val GTT	;
	. XX	i leu	
	10 K	1 1y3	
25	2 91°	1 AV	}
	\$ 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8	1 9 <sup>1</sup> )	
	0 1 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4	o gle	
	540 a thr	570 9 glu 3 GAG	
20		r GCC	
30	A 1.	s phe	
	ى ئى ئى كى	56. 1 Cys. 1 TGA	
	ه د ک	glu thr GAG ACC ter ter	
	leu val lys his lys pro lys CTC GTG AAA CAC AAG CCC AAG	561 ale asp asp lys glu thr cys phe GCT GAC GAT AAG GAG ACC TGC TTT ter ter CATCTCAGCCTACCATGAGAATAAGAGAAAGA	
35		7 AAC	
	ر د و و و و	S CAS	
	0 c -	1 88 T 27.7.7.7.7.7.7.7.7.7.7.7.7.7.7.7.7.7.7.	
	531 91u 6AG	3 5 Z	

- 10. A nucleotide sequence according to any of claims 6 to 9, in essentially pure form.
- 11. A DNA transfer vector comprising a nucleotide sequence as defined in claim 5.
- 5 12. A DNA transfer vector according to claim 11, transferred to and replicated in a micro-organism.
  - 13. A DNA transfer vector according to claim 12, which is a plasmid.
- 14. A DNA transfer vector according to claim 13,
- 10 wherein the plasmid is pBR322 or YEp6.
  - 15. A process for preparing human serum albumin, which comprises culturing a micro-organism according to claim 5.
  - 16. A DNA transfer vector according to any of
- 15 claims 12 to 14, or a process according to claim 15, wherein the micro-organism is a bacterium or yeast.
  - 17. A vector or process according to claim 16, wherein the bacterium or yeast is <u>E. coli</u> or <u>Saccharomyces cerevisiae</u>.

